L Number	Hits	Search Text	DB	Time stamp
1	7660	(chem$7 or substance$3) adj (library or libraries)	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/09/17 20:38
2	5072	((chem$7 or substance$3) adj (library or libraries)) and parameter$5	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/09/17 20:39
3	2	((chem$7 or substance$3) adj (library or libraries)) and (production adj parameter$5)	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/09/17 20:47
4	13	((chem$7 or substance$3) adj (library or libraries)) and (test$3 adj parameter$5)	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/09/17 20:47
5	1873	702/22-32.ccls.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/09/17 20:52
6	117	702/22-32.ccls. and @pd>=20040517	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/09/17 20:53
7	28	(702/22-32.ccls. and @pd>=20040517) and (library or libraries)	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/09/17 20:53

Point: The manner in which the information is organized, generated, analyzed, and used.

1. Get alpha version of software co-designed, initially on combinatorial chemistry, later on all synthetic chemistry and structural biology.

2. Use software plus a team of scientists to design initially libraries and later more complicated disease target solutions.

3. Generate a range of human and computer solutions to the problems using a network of individual minds and an artificial intelligence neural network with both groups working with the same information and feedback loops to evaluate and improve all methods.

4. Share the information, with a short delay, with the rest of the scientists over the internet. Keep the code for how the information was generated and analyzed internal.

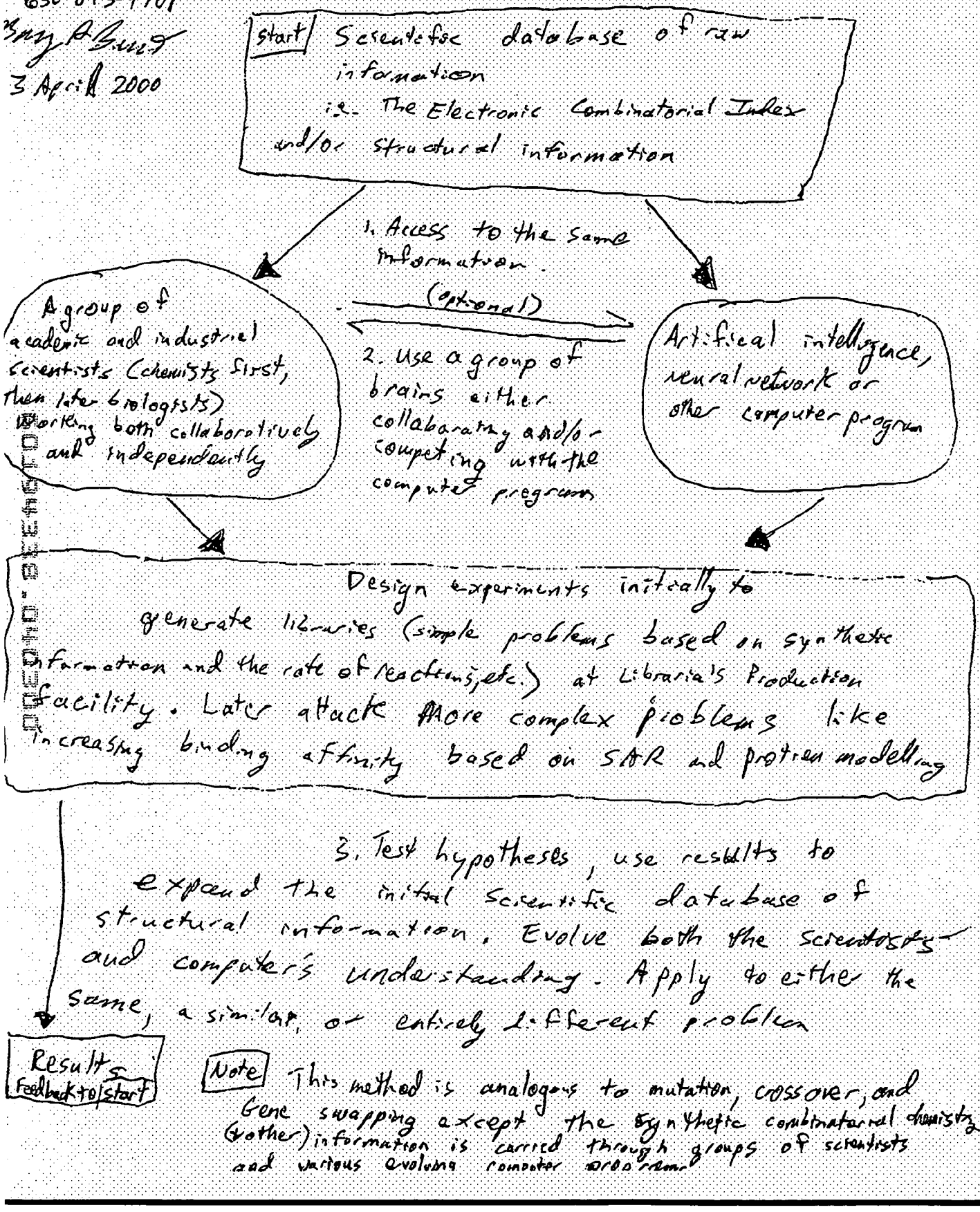
5. Evolve the database, software, and human team as appropriate for each problem. This acknowledges that there will be no general single solution and provide the impetus testing new models in a Darwinian Sense.

6. Put in checkpoints, like Asimovs Robots rules of order to keep the system appropriately focused.

2 April 2000 Barr Bauer.

Dr. Barry A. Bunin
bunin@combinatorial.com
650-873-7704

Barry A Bunin
3 April 2000



Note: This method is analogous to mutation, crossover, and Gene swapping except the synthetic combinatorial chemistry (other) information is carried through groups of scientists and various evolving computer programs.

Libraria's Broad Patent Application(s) Raw Information

Patent application(s) tentative title:
"Tying an Evolving Database of Chemical Information to Flexible Services"

I. Introduction, Purpose, and Scope:
- A. The Purpose of this patent is to cover (but is not limited to):
  1. The organization of the information
  2. The search of the information
  3. The evolution of the information
  4. Tying chemical services to the information
- B. The field of "Combinatorial Chemistry"
  1. History and relationship to all synthesis
  2. Current state of the art
  3. How tying flexible services to "The Electronic Combinatorial Index" format represents a novel, non-obvious product(s)
- C. The field of "Artificial Intelligence"
  1. Previous applications to chemistry (ironically the Dendral project at Stanford applying AI to molecular structural identification based on mass spectroscopy data was one of the first expert systems published)
  2. Expert Systems
     a. Creation of databases
     b. Agents to search databases
     c. Recent advances (including web-industry technology, multi-agent teams, the human/computer interface, and inductive learning in flexible hypothesis space)
- D. The proposal is to build a system that grows information databases and searches information databases with a team of scientists and artificial intelligence capabilities starting with combinatorial chemistry (the most general and reliable synthetic methods) and having feedback loops that help both groups learn. At every stage when practical the information system will be tied to real services (for example, the synthesis of libraries and individual compounds).
- E. Network approaches and open systems for information and services
- F. Possibility for multiple patents:
  1. Patents related to the organization of chemical (and other) information
  2. Patents related to the Evolution of the organization of chemical (and other) information
  3. Patents on the search capabilities of the chemical database
  4. Patents related to the organization of precursors and selection (i.e. software, sets, etc.)
  5. Patents on the methods for tying flexible services to these patents

II. Building a scaleable architecture for synthetic and related databases
   A. Current efforts with static databases
   B. Our potential architecture described in gory detail (while acknowledging the final product(s) will change, imagine each inner circle/database is a subset of the next larger circle graphically like a bullseye for darts):
      1. Central position is "The Electronic Combinatorial Index"
      2. Second circle includes widely used chemistry texts and reference works
      3. Third circle includes leading references from the second circle
      4. Fourth circle includes leading journals
      5. Fifth circle includes chemical databases that are commercially available
   C. Create a simple, uniform format with actual procedures and reliability ratings that are simple enough for most humans to easily use yet robust enough for an artificial intelligence computer program to work with and evolve additional tasks, information, and networks
   D. Detailed description of a system that can evolve to include and work with:
      1. Related fields such as structural biology and drug discovery
      2. Unrelated fields such as economics and psychology database
   E. Examples of evolved utility with open systems
      1. Sciquest and Chemdex offerings
      2. Web search engines
      3. Prous science
      4. Dendral and meta-Dendral
      5. Aldrich chemical catalog
      6. Cambridge software's web portal

III. Detailed description for tying chemical services to the information databases

IV. Network approach with open systems to develop:
   A. New applicable chemistries
   B. New AI applicable technologies
   C. Unique groups of experts to interact with the evolving database of information and to suggest improvements/additions after each use (perhaps for a discounted price or similar benefit, etc.).

I. Introduction, Purpose, and Scope:

A. The Purpose of this patent is to cover (but is not limited to):

The field of synthesis has evolved from observations and experiments of scientists for over a century. The information has been documented in the primary literature as well as a number of reference texts on synthetic methods. More recently combinatorial chemistry has rapidly developed as a subset of all synthetic chemistry particularly in the last decade. Combinatorial chemistry can be though of as the most general and reliable methods for high-throughput synthesis. It often involves automation and has been likened to the industrial revolution finally meeting synthetic chemistry. The technology has increased the ability of a chemist to rapidly elucidate structure-activity relationships but has often been under exploited for a number of reasons. Primary roadblocks include the lack of sufficient information on reliable methods for library generation and the difficulties associated with optimizing reaction conditions to provide the desired products in high yield. These patent(s) describe how one can create, organize, search, evolve and actually use databases of synthetic information in conjunction with combinatorial chemistry services.

B. Summary of Combinatorial Chemistry:

Note: The following pages as background material are from the early chapters of my widely used text on combinatorial chemistry "The Combinatorial Index".

"The goal of this book is to provide practical information about the combinatorial synthesis of small molecules for a range of individuals interested in the field, including bench chemists, lab managers, medicinal chemists, agrochemists, academic chemists, computational chemists, and even biologists. Regardless of one's experience in combinatorial chemistry, many questions inevitably arise, often simple questions such as: What has already been done? How was it done? Where can the necessary materials or instruments be found? With an explosion in information, the organization of the information becomes crucial. The Combinatorial Index provides a compilation of the synthetic information for constructing combinatorial libraries of small molecules.

The synthesis and screening of libraries of nonbiological oligomers (see Appendix 1) and nonpolymeric organic compounds ("small molecules") have rapidly become the focus of intensive research efforts. The Combinatorial Index describes the methods reported in the literature for the rapid preparation of functionally diverse small molecules. Points of interest, representative examples, and literature procedures summarize each method. In addition to focusing on small molecules, The Combinatorial Index is a compilation derived exclusively from journal articles (i.e., patent literature is not included).[0] Comments on the original studies are made; however, the majority of the information presented is factual and is thus reported as described in the primary literature. In cases where the copywritten material could not be obtained, the procedures were paraphrased as literature summaries. References are provided at the beginning of each new heading. Although this compilation provides an overview of the combinatorial synthesis of small molecules, it is primarily a resource book.

Many combinatorial libraries are currently constructed using solid-phase synthesis. Therefore, following a brief introduction and background, three chapters are devoted to the different linkers, reactions, and analytical techniques for solid-phase synthesis. Although many reactions on support are high yielding, significant optimization is often required before they are efficient and general enough to be used to construct combinatorial libraries. It is hoped that providing examples of different reactions along with the specific conditions that were necessary for optimization will assist related combinatorial studies. Representative examples are given to help assess the generality and limitations of the different methods. Reliable methods for reactions such as substitutions, cyclizations, condensations, and Suzuki couplings are included specifically because they can be used in different contexts.

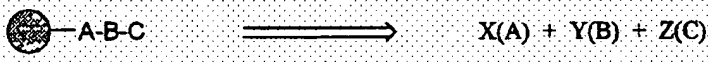
Doing organic chemistry on solid support has been likened to working with a blindfold on because of the limited analytical and purification techniques available relative to those available in solution. This is one of the arguments in favor of building libraries in solution. The most direct way to

evaluate the fate of a particular set of reaction conditions on support is to cleave the material off of support and rigorously characterize the products. Unfortunately, this is not always possible because intermediates are often unstable to cleavage conditions. Furthermore, particularly in a multistep sequence, there are often faster methods for determining whether a particular reaction worked (i.e., Fmoc quantitation). Detailed Procedure for a range of different quantitative and qualitative analytical methods for solid-phase synthesis are described in chapter 5.

A growing number of reports on solution libraries have appeared in the literature. Criteria and methods for building solution libraries are discussed in Chapter 6. Whether a library is prepared on support or in solution is often dictated by the type of chemistry being developed (or vice versa, the type of chemistry being developed is often dictated by whether a library is prepared on support or in solution). High-throughput purification techniques such as solid-phase and liquid-phase extractions are often critical for preparing solution libraries that are useful for screening. The challenges associated with the construction of solution libraries (e.g., solubility and purification) can be quite different from the challenges associated with solid-phase combinatorial synthesis (e.g., monitoring reaction progress and scale up). Although there are many differences between the solution-phase and solid-phase strategies for generating libraries, in both cases the synthetic challenge is to develop reaction conditions that are general and high yielding.

Often I have chosen to be inclusive, rather than selective, in citing examples from the literature. Even so, due to the large body of literature, many examples are cited in related studies. For example, in the sections discussing subjects such as the formation of amides and esters, only selected examples are provided for obvious reasons. I have made every effort to be both fair and informative about the strengths and weaknesses of the various methods in this compilation. I apologize for any topics that were underrepresented or misrepresented. I would appreciate being notified at bunin@combinatorial.com of new or incomplete information for incorporation into future editions or online supplements.

A salient feature of combinatorial synthesis is that a large amount of diversity can be generated from a relatively small number of building blocks. A representative example of a simple combinatorial library prepared on solid support from three sets of building blocks, A, B, and C, is illustrated below. From only 10 derivatives of each



X + Y + Z = Total number of variables used as inputs in the library

(X)(Y)(Z) = Total number of compounds generated from the library synthes

building block, a library of 1000 trimers can be generated; with 100 derivatives of each building block, 1,000,000 compounds can be accessed. With rapid access to such large numbers of compounds, new issues arise such as which compounds are the most useful to make and how to keep track of the large amount of information that is generated.

Currently, there are a number of distinct approaches for generating combinatorial libraries *in vitro*. The compounds can be synthesized in a spatially separate format or as pooled mixtures. A number of methods for identifying active compounds in a mixture have been developed. Obviously, identifying an active compound is straightforward when the compounds are synthesized in a spatially separate format. A brief overview of the different methods for preparing synthetic combinatorial libraries follows. More detailed discussions can be found in a number of review articles.[1, 2, 3, 4]

Methods for Generating Combinatorial Libraries.

A. Spatially Separate Synthesis.

The most straightforward approach for library analysis is to keep the different compounds (or other variables) spatially separate in a parallel array. The primary advantage of keeping the compounds spatially separate is that it removes some of the ambiguities associated with pooling compounds. When the compounds are spatially separate, direct structure–activity relationships are obtained from biological

evaluation. Analytical evaluation of the chemical integrity of the compounds is also straightforward when the compounds are spatially separate. The primary disadvantage of spatially separate libraries is that the number of compounds that can be synthesized is more limited.

The first combinatorial library was prepared in a spatially separate format by Geysen and co-workers in 1984.[5] They developed functionalized pins for solid-phase peptide synthesis and epitope analysis. The pins were configured to be compatible with 96-well microtiter plates. The pin technology has been improved using different polymers, as well as higher loading levels and functional linkers to accommodate other chemical applications.[6] Fodor and co-workers at Affymax have developed photolithographic methods for building large libraries on a silicon wafer.[7] Large spatially separate libraries (100,000 compounds) can be prepared with this method. However, because it requires photolabile protecting groups and support-bound biological assays, the technology is primarily being applied to DNA diagnostic tests.[8] A number of new technologies for the preparation of spatially separate libraries on resin and in solution are currently being developed.

B. Split Synthesis. There are a number of different pooling strategies. The earliest of these, developed independently by Furka,[9] Lam,[10] and Houghten,[11] employ a split and mix procedure to generate mixtures of peptides. In a split synthesis, a quantity of resin is split into equal-sized portions in separate reaction vessels and reacted with different monomers. After the reactions are complete, the resin is pooled together and thoroughly mixed. A common protecting group can be removed, or a common transformation can be performed, in a single reaction vessel. For the coupling of a second monomer, the resin is split again, and the process is repeated until the end of the combinatorial synthesis. To couple different building blocks, such as activated amino acids, the resin must be split into separate reaction vessels to allow reactions with different rates to be driven to completion.[12]

There are a number of techniques for identifying biologically active components from a library prepared by a split synthesis. The active components in a mixture can be isolated by deconvolution studies such as an iterative resynthesis and evaluation of smaller pools. A portion of the resin can be saved at each step to facilitate the iterative resynthesis. In addition, orthogonal,[13] positional,[14] and indexed[15] libraries all use pooling strategies that minimize the amount of deconvolution required.

The combinatorial methods initially developed for peptide synthesis have also been applied to the combinatorial synthesis of unnatural biopolymers and small molecules. In one early example, high-affinity ligands to 7-transmembrane G-protein-coupled receptors (7TM/GPCR) were identified from the split synthesis of a diverse peptoid library.[16]

At the end of any split synthesis, each individual bead theoretically contains a single product, since all of the sites on any particular bead have been exposed to the same synthetic reagents. "One-compound, one-bead" approaches have been developed to identify the active components in a biological assay without resorting to a time-consuming iterative resynthesis. With certain assays of support-bound compounds, an active compound from a single resin bead is identified after it binds with a radiolabeled or fluorescent-labeled receptor.[17] After active components on support are detected and isolated, the chemical structure can be determined using a method such as Edman degradation for the identification of support-bound peptides. Methods for the partial release of compounds off the support have been developed for biological evaluation in solution. After biological evaluation, the compound that remains on the resin beads can be used for structural identification.[18]

A conceptually different approach to deconvoluting active components from a library prepared by split synthesis involves a molecular tagging scheme. In this approach, readable tags that encode the reaction sequence are attached to resin. DNA was an obvious choice for encoding,[19] since that is what Nature uses. Unfortunately, DNA is not chemically stable under many of the reaction conditions frequently used in organic synthesis. To circumvent this problem, encoding has been performed with peptides prepared from amino acids that have relatively unreactive side chains[17] or GC–EC tags that are inert to most of the reaction conditions typically employed.[20] The advantages of the GC--EC tags, developed by Still and co-workers, are that they can be both detected at less than 0.1 pmol and attached directly to polystyrene via carbene chemistry. Thus, the method does not require an orthogonal protecting strategy. Radiofrequency tagging strategies have also been developed as an alternative method for encoding libraries on resin.[21,22] Alternative approaches to generating combinatorial libraries and optimizing biological activity, such as genetic algorithms, are currently being investigated.[23,24]

At least as important as the format in which libraries are prepared are the classes of compounds that are accessible. This compilation describes synthetic methods and analytical techniques to assist in the development of chemistry for combinatorial libraries." (from Bunin, B. A. The Combinatorial Index, Ch. 1-2, 1998, Academic Press).

By expanding the book to a suite of software products and services their is an opportunity to create rapidly accelerate drug discovery in a way that currently non-obvious to the ordinary chemist. A key component to this strategy will be the ability to expand the initial database on combinatorial chemistry to incorporate all synthetic chemistry. Another key component will be to incorporate flexible services as part of the software package. The way in which the database will evolve is the final critical component. This will tap heavily into related emerging fields of artificial intelligence.

C. The field of "Artificial Intelligence"

The field of Artificial Intelligence roots can be traced to the now famous Turing test for computer intelligence.[25] The basic postulate is that rather than ask if computers can think, the more testable question is given a series of questions can an interrogator determine if the typewritten answers are coming from a human or a computer. A wealth early studies in the field can be found in the classic book Computers and Thoughts for more detailed background.[26]

Ironically, one of the first applications (expert systems) to use AI was the Dendral project that assisted with molecular structure identification based on mass spectroscopy data.[27, 28, 29] In the Dendral program was a collaboration initiated between Feigenbaum, Lederberg, Buchanan, and Djerassi to elucidate chemical structure at a high level of competence. Given a molecular formula, the spectrographic data, and encoded heuristic knowledge of organic chemists, the Dendral interactive program explores possible molecular configurations in the search for the true structure. The project helped elucidate some of the basic mechanisms of hypothesis generation and evaluation. The results of the project suggested that knowledge was as important as reasoning in these systems. In any case, today there are many examples where artificial intelligence has been used to generate expert systems with varying degrees of success.

Expert systems attempt to replicate the decision making process of a human expert in a limited field. It consists of three components: a knowledge base, decision rules, and an inference engine.[30] As with all developing technologies, without an appropriate problem they are academic endeavors. The real utility involves how a technology (in this case evolving technologies of artificial intelligence and combinatorial chemistry) is applied to a specific problem. We have a unique set of interrelated problems that can be solved with these technologies leading to efficiencies to society, initially in the field of drug discovery, that are not currently obvious.

There have been a number of current advances that allow us to create products not previously available. In particular, because of the nature of the problem we can time the development of future products with the development of both growing fields. A few examples of current advances relevant to the problems include but are not limited to web-industry technology,[31] multi-agent teams,[32] the human/computer interfaces,[33] and inductive learning in flexible hypthesis space.[34] These trends have implications for overcoming the challenges to feedback-directed optimization in software database development. Current advances in the reasoning component will receive human checks and balances in the development of a new types of scientific databases linked to services. This...

D. The Proposal:

To build a system that grows information databases and searches information databases with a team of scientists and artificial intelligence capabilities starting with

combinatorial chemistry (the most general and reliable synthetic methods) and having feedback loops that help both groups learn. To maintain relevance at each step the information must be evaluated and corrected by humans and computer programs until it is evident which is better suited for which specific tasks. This could be done with crossover of information so both teams collaborate and compete. At every stage when practical the information system will be tied to real services (for example, the synthesis of libraries and individual compounds).

E. Network approaches and open systems for information and services

In the long run, the key to expanding the systems will be to make it open and compatible with outside parties. This approach is applicable to developing the database, developing the search engines, developing the basic technologies (both AI and chemical), and interacting with other databases. The epitome of this approach is the attraction of the Internet. In our case, there will be third party contributions to content, basic research, and software architecture. At a certain point there will be a positive spiral that creates uniquely evolving database information products (initially in chemistry). Part of the architecture of the suite of products will be a filter of suggestions and additional contributions.

F. The possibility for multiple patents:

Although the stated purpose of this document is "to build a system that grows information databases and searches information databases with a team of scientists and artificial intelligence capabilities starting with combinatorial chemistry (the most general and reliable synthetic methods) and having feedback loops that help both groups learn. At every stage when practical the information system will be tied to real services (for example, the synthesis of libraries and individual compounds)." This does not in any way limit the potential claims and future applications. The information contained in this document set the foundation for multiple possible patents, including but not limited to:

1. Patents related to the organization of chemical (and other) information
2. Patents related to the evolution of the organization of chemical (and other) information
3. Patents on the search capabilities of the chemical database
4. Patents related to the organization of precursors and selection (i.e. software, sets, etc.)
5. Patents on the methods for tying flexible services to these patents

These and other applications are described in detail in the following sections.

II. Building a scaleable architecture for synthetic and related databases

A. Current efforts with static databases

A book is a static database. It contains information, but has no ability to evolve its internal structure once the book is printed. An intermediate level is a software database that provides the user with a number of options to choose from and a number of possible answers to queries. A more advanced expert system would continuously evolve based on feedback loops. The Electronic Combinatorial Index database can evolve into something much greater than the initial static product. Procedures for combinatorial chemistry represent a subset of all synthesis, procedures for synthesis represent a cross section of procedures for drug discovery, procedures for drug discovery are a cross section of all chemistry and biology. The details of when and how the connections are made are critical to the growth of the product(s). The architecture will contain procedures (information) and deliverables (services) at multiple stages. The internal connections will become stronger as they are used

and the number of external connections will increase much like the development of an embryonic brain.

B. Our structure for evolving databases:

The central position will be a set of reliable methods for high-throughput combinatorial synthesis. This is analogous to the relationship combinatorial synthesis has to all synthetic methods. As previously mentioned, combinatorial synthesis represents the most general, expedient and robust synthetic methods because by their very nature they should be tolerant of a range of functional groups. While this central position will inevitably grow over time as additional scientists publish on combinatorial methods, the real growth in the database will be a result of the tentacles that reach into the more mature field of chemical synthesis as rather broadly defined.

The second position will include the most widely used chemistry referenced books judiciously selected. The procedures from the leading references will be included along with their lineage to the more general high-throughput synthesis methods. Examples of appropriate reference works for the second circle include, but are not limited to March's Organic Synthesis text, Green and Wuts' Protecting Groups in Organic Synthesis, Bodansky and Bodanskys' Practice of Peptide Synthesis, The Encyclopedia of Organic Reagents. All of the facts and information from texts will be reformatted in a uniform simple manner for easy access by humans and computer programs and to avoid any copyright restrictions. Other reference texts that will be added to the second circle, include but are not limited to Organic Synthesis, Organic Reaction, Org. Syn Prep., and Comprehensive Heterocyclic Chemistry.

The third circle will include the leading references from the first two circles. When appropriate, this will lead to a chain of articles as commonly done when an individual learns about a new field while spending time in a chemistry library.

The fourth circle will actually start to include a systematic reorganization of the entire chemical literature in the journals. Reorganizing the information in a common format will have obvious advantages for the end user. This is similar to chem abstracts except that it will emphasize procedures, tye them to services, and make them part of an evolving database.

Once the inner circles reach a certain mass inertia it will be time to incorporate other databases. This is why it is critical to develop very robust search agents at the early stage of the project that are able to search other databases and vice-versa. The key to utility will be in the organization of the material and the simplicity (with enormous flexibility) of the search engine.

The methods for organization, representation, and uses of databases described herein should be applicable to both related and unrelated fields of knowledge.

C. Representation of information:

It will be important to create a simple, uniform format with actual procedures and reliability ratings that are simple enough for most humans to easily use yet robust enough for an artificial intelligence computer program to work with and evolve additional tasks, information, and networks.

D. Detailed description of a system that can evolve to include and work with:

1. Related fields such as structural biology and drug discovery.
2. Unrelated fields such as economics and psychology database.

We are creating a chemo-informatics tool that is as easy and simple to use while still being scaleable and tied to services whenever possible. It will be an appropriate real life

valuable problem to evolve various competing artificial intelligence technologies and approaches. It will be an open system that is compatible with other databases and information searching methods. It will include agents to search for information that are guided by both an artificial intelligence program and a 24 hour available human expert in chemical database mining.

The chemo-informatics tool includes the most valuable information such as synthetic procedures with reliability ratings based on experimental data. All information will be one click away from a "shared risk" feasibility study and custom services. The services include but are not limited to the delivery of individual compounds, small libraries (bookends of 10-100), and large libraries (1000-100,000). With the right infrastructure, systems, and people we will routinely deliver custom services faster.

With our general diversity and targeted libraries we will include first generation actual and virtual libraries. The libraries will come with software for SAR trees describing follow-on libraries. This will include the option of using other software and biasing the SAR trees describing follow-on libraries with known structural information. These will be tools that can be used in any laboratory around the world. For example, each set of 2000 compounds (2D library) and 6000 compounds (3D library) on average will come with an option for a second generation library based on the SAR or a full blown-out library (perhaps in a split and mix format) if the customer wishes to hide the SAR. Both human and AI teams would collaborate and compete on the reaction condition optimization and SAR optimization problems.

The informatics package will be linked to a network of cross references to suggest related sets of compounds for synthesis. For example, a section from March's Organic Chemistry text might lead to series of papers that suggest a particular set of reaction conditions. Both the computer and humans will find patterns of "leading references". The key will be, over time, to include reliability ratings. The data output should include references, words, and schemes as provided by current chemical databases such as Bielstien while including addition outputs such as procedures and the potential to have a compound described in the literature made for you and delivered (following shared-risk feasibility). The customer would have the option of using our information, using our synthetic services, or a combination of the two offerings.

The organization based services would be initially targeted at the synthetic chemist. The value of the services would be proportional to the time, scale, and difficulty. Offerings would include but are not limited to "one step from Aldrich, three steps from Lancaster, two steps in Bielstien" for example. Reliability ratings will help with assessing difficulty.

A key component will be to have the structural and reaction infomration stored in a uniform manner so both chemists and AI programs get used to improving the network and making connections. Each time a human expert finds a new connection (s)he will notify the computer program and vice versa. The network technology will evolve on the coattails of web technology currently being paid for by Yahoo, Inkotomi, Googgle, and others in addition to in-house expertise.

The information and database services could be extended to include analytical data (theoretical and/or experimental), structural data on the molecules, biostructural data for more complex problems, chemical ordering information, and web-offerings. These are all related fields under the umbrella of drug discovery.

The information and database service could be extended to partially related fields like polymer synthesis and enzymology. A truly robust system could also be extended to unrelated fields. This approach could be used, with the judicious selection of starting points and experts, to organize medicine in a standard and simple format. The approach could be applied to any other field where significant bodies of information our guided by underlying logical principles. Obvious examples include psychology, law, engineering, archetecture, journalism, economics, history, business, electronics, and the internet to mention some possibilities. To do this practically, since we will be late players in these fields, will require

a strong technology base using the state of the art in both human experts and artificial intelligence.

The guiding principles would be the same as those used in organizing the synthetic information. Namely, identify the most valuable content (such as synthetic procedures) and develop a uniform method of representation. For example, instead of randomly surfing the Internet, use peer reviewed content from scientific journals and books. We will carefully select the content to develop intuitive products. Finally, we will continue to offer services either by building in-house expertise or partnering if and when appropriate.

E. Examples of evolved utility with open systems

There are a number of offerings that have demonstrated the utility of open systems that leverage the efforts of others. In chemistry, Sciquest, Chemdex, Cambridge Soft's web portals are more recent examples. Historically, the Aldrich chemical catalog is a traditional example whereby chemicals are bought elsewhere and rebottled or made in house with the same result for the customer. To the extent possible, the Libraria flexible service offering will take into account a range of requests and then search the world's database of synthetic information for appropriate solutions.

III. Detailed description for tying chemical services to the information databases

Because we will be reorganizing the chemical information as it is added to the database, although it will be an ambitious task but we will be able to pick and choose what to include and to assign reliability ratings (either based on theoretical or actually experimental data). Some of this information will be obtained from third party laboratories.

Related to the concept of creating a uniquely consistent format for organizing chemical information is the ability to develop a system to tie a flexible range of services that are not currently possible to this information. As one example, instead of having a follow-on library with just the same reactions (which is the most straightforward approach both scientifically and logistically) have a software program(s) and/or human(s) select from a range of chemistries with reliability ratings either as diverse or targeted sets of library products and services.

Another way to tie information to services is to organize flexible sets of precursors based on structural criteria in the literature (see, Lipinski, C. et al and Murcko, M. et al). This will be organized in a uniform, yet uniquely flexible organization of the data input variables for library generation (both for individual libraries or groups of libraries as described in the first example). The precursors can be organized and barcoded in rows and columns of a grid such as the industry standard 96-well microtitor plate. Additional information could be added to the precursor sets piror to internal or external use including but not limited to solubility data, reliability ratings (+/- or 1-10), aromatic/aliphatic, diverse/targeted, hydrophobic/hydrophilic, alpha/beta substituted, o,m,p-substituted, ring size, bicyclic/fused, etc. Customers would have the option of using the software tools or their own in the selection process. They could test the chemistry in their own laboratory (one place) with a subset of precursors and then generate the full library in our laboratory (second place) because of the careful preorganiation and preselection of appropriate data input precursors. These preorganization of these sets of precursors would facilitate "shared risk" rapid feasibility studies on precursor compatibility with new chemistries. Finally, it would facilitate library design, feasibility, and generation functions.

The preorganization of both synthetic information and structurally relevant precursor sets will allow for more flexible chemical services. These could range from individual compound synthesis, to small sets (bookends), to large sets (full libraries). The key will be to build up a critical mass of intelligently organized, flexible chemical information (both in computers and people) to offer the greatest range of services. The guiding principles could be applied to other related and unrelated fields.

I. Network approach with open systems to develop:
   A. New applicable chemistries
   B. New AI applicable technologies

C. Unique groups of experts to interact with the evolving database of information and to suggest improvements/additions after each use (perhaps for a discounted price or similar benefit, etc.).

To truly realize the full potential of this approach would require a substantial effort to create industry standards with open systems and then to find ways to work with other experts in their various fields. For chemistries this would include experts from acadamia as well as combinatorial chemistry laboratories. A similar network could help evolve the artificial intelligence applications and technologies once a robust framework for programming is created. Finally, the extent to which the system tracks use and has feedback loops based on standard representations of data is the extent to which the system can evolve.

List of possible alliances:

For combichem:
Axys
DPI
Trega
ArQule
Cambridge Combinatorial
Molecumetics
NCE
3D Pharmaceuticals
Chemcodes
UC Berkeley
Harvard
Stanford

For webstrategy:
ChemNavigator
ChemRoutes

For software development:
Scivision
Tripos
Stanford
U of T, Austin

For web logistics:
Sciquest
Chemdex
Cambridge Soft

For building blocks:
Aldrich
Argonaut
Lancaster

For instrumentation:
Robbins/Genevac
Hewlett-Packard
Berger Instruments

Big Pharma (especially Merck)

Contact:
Dr. Barry A. Bunin
Tel 650-873-7704
Fax 650-873-7707
bunin@combinatorial.com

---

[0]. For a detailed review of recent solid-phase organic reactions including the patent literature, see Hermkens, P. H. H.; Ottenheijm, H. C. J.; Rees, D. *Tetrahedron* **1996**, *52*, 4527–4554.

[1]. (a) Thompson, L. A.; Ellman, J. A. *Chem. Rev.* **1996**, *96*, 555–600. (b) Choong, I. C.; Ellman, J. A. *Annu. Rep. Med. Chem.* **1996**, *31*, 309--318.

[2]. (a) Gallop, M. A.; Barrett, R. W.; Dower, W.; Fodor, S. P. A.; Gordon, E. M. *J. Med. Chem.* **1994**, *37*, 1233–1251. (b) Gallop, M. A.; Barrett, R. W.; Dower, W.; Fodor, S. P. A.; Gordon, E. M. *J. Med. Chem.* **1994**, *37*, 1385--1401.

[3]. (a) Jung, G.; Becksickinger, A. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 367–383. (b) Früchtel, J. S.; Jung, G. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 17–42.

[4]. Rinnova, M.; Lebl, M. *Collect. Czech. Chem. Commun.* **1996**, *61*, 171--231.

[5]. Geysen, H. M.; Meloen, R. H. Barteling, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 3998–4002.

[6]. Maeji, N. J.; Valerio, R. M.; Bray, A. M.; Campbell, R. A.; Geysen, H. M. *React. Polym.* **1994**, *22*, 203–212.

[7]. Fodor, S. P. A.; Read, J. L.; Pirrung, M. C.; Stryer, L.; Lu, A. T.; Solas, D. *Science* **1991**, *251*, 767–773.

[8]. Abbott, A. *Nature* **1996**, *379*, 392.

[9]. Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 487–493.

[10]. Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature* **1991**, *354*, 82–84.

[11]. Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. *Nature* **1991**, *354*, 84–86.

[12]. Houghten has found that the relative rates of amide bond formation on solid support with different amino acids were primarily dependent on the activated amino acid in solution. By adjusting the concentration of the amino acids to compensate for their relative rates, different amino acids can be pooled in the same reaction vessel.

[13]. Deprez, B.; Williard, X.; Bouirel, L.; Coste, H.; Hyafil, F.; Tartar, A. *J. Am. Chem. Soc.* **1995,** *117,* 5405--5406.

[14]. Pinilla, C.; Appel, J. R.; Blanc, P.; Houghten, R. A. *Biotechniques* **1992,** *13,* 901--905.

[15]. Pirrung, M. C.; Chen, J. *J. Am. Chem. Soc.* **1995,** *117,* 1240--1245.

[16]. Zuckermann, R. N.; Martin, E. J.; Spellmeyer, D. C.; Stauber, G. B.; Shoemaker, K. R.; Kerr, J. M.; Figliozzi, G. M.; Goff, D. A.; Siani, M. A.; Simon, R. J.; Banville, S. C.; Brown, E. G.; Wang, L.; Richter, L. S.; Moos, W. H. *J. Med. Chem.* **1994,** *34,* 2678--2685.

[17]. Vetter, D.; Tate, E. M.; Gallop, M. A. *Bioconjugate Chem.* **1995,** *6,* 319--322.

[18]. Lebl, M.; Patek, M.; Kocis, P.; Krchňák, V.; Hruby, V. J.; Salmon, S. E.; Lam, K. S. *Int. J. Pept. Protein Res.* **1993,** *41,* 201--203.

[19]. Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993,** *90,* 10700--10704.

[20]. Ohlmeyer, M. H.; Swanson, R. N.; Dillard, L. W.; Reader, J. C.; Asouline, G.; Kobayashi, R.; Wigler, M.; Still, W. C. *Proc. Natl. Acad. Sci. U.S.A.* **1993,** *90,* 10922--10926.

[21]. Moran, E. J.; Sarshar, S.; Cargill, J. F.; Shahbaz, M. M.; Lio, A.; Mjalli, A. M. M.; Armstrong, R. W. *J. Am. Chem. Soc.* **1995,** *117,* 10787--10788.

[22]. Nicolaou, K. C.; Xiao, X. Y.; Parandoosh, Z.; Senyei, A.; Nova, M. P. *Angew. Chem., Int. Ed. Engl.* **1995,** *34,* 2289--2291.

[23]. Weber, L.; Wallbaum, S.; Broger, C.; Gubernator, K. *Angew. Chem., Int. Ed. Engl.* **1995,** *1,* 2280--2282.

[24]. Sheridan, R. P.; Kearsley, S. K.; *J. Chem. Inf. Comput. Sci.* **1995,** *35,* 310--320.

[25]. Turing, A. M. *Mind,* **1950,** *59,* 433-460.

[26]. Feigenbaum, E. A.; Feldman, J. "Computers and Thought" AAAI Press Edition, 1995, Menlo Park, first published in 1963 by McGraw-Hill Book Company.

[27]. Buchanan, B. G.; Feigenbaum, E. A., *Artificial Intelligence,* **1978,** *11,* 5-24.

[28]. Lederberg, J. "How DENDRAL Was Conceived and Born", ACM Symposium on the History of Medical Informatics, National Library of Medicine, 1987.

[29]. Lindsay, R. K.; "Applications of Artificial Intelligence for Organic Chemistry: The DENDRAL project", McGraw-Hill Publishing, 1980.

[30]. Kurzweil, R. "The Age of Intelligent Machines", p 292 The MIT Press.

[31]. MIT CBCL and AI lab cosponsered "Spring 2000 Brains and Machines Seminar Series" February 23 Tom Mitchell "Supervised Learning from Unlabeled Data". On the the internet at http://www.ai.mit.edu/events/brainsMachines/abstracts/S2000/mitchell

[32]. Kaminka, G. A.; Tambe, M. *J. Artificial Intelligence* **2000,** *12,* 105-147. On the internet at http://www.cs.washington.edu/research/jair/abstracts/kaminka00a.html.

[33]. "Human Factors in Information Systems: An Organizational Perspective" Ed. Carey, J. M. *Ablex Publishing Corp.* 1991, Norwood, NJ.

[34]. Baxter, J. "A Model of Inductive Bias Learning", *J. Artificial Intelligence* **2000,** *12,* 149-198.